FLIGHTS DATAWAREHOUSE AND SYSTEM ANALYTICS

Presented by

- AYAN ADNAN (DT-22014)
- MUHAMMAD AMAN QAZI (DT-22044)
- SYED AWAIS WASEEM (DT-22033)
- AAMIR RASHID (DT-22047)

COURSE CODE: CT-472

COURSE INSTRUCTOR: Dr. Muhammad Umer Farooq

Introduction

This presentation outlines the development of a unified analytical platform designed to centralize airline operational data. Our primary goal is to integrate diverse data sources—OLTP databases, real-time APIs, and customer feedback—into a single, structured repository built using Python, SQL Server, and Power BI, adhering to ETL best practices.

Project Objectives

Centralized Data Storage
- Seamlessly combine multiple data sources including OLTP databases, REST APIs, and CSV files into a cohesive unified system for holistic airline operations view.

Business Intelligence Enablement
- Fast querying & visualization ready

Data Quality & Consistency
- Clean, validated, standardized data

Automation & Scalability
- Automated scheduling with Windows Task Scheduler
- Continuous updates & future growth support

Business Intelligence
- Interactive Power BI dashboards for insights

Data Sources Overview

01.OLTP Database 1

- Tables: Customers, Bookings, Payments
- Content: Customer demographics, booking transactions, payment records
- Purpose: Core transactional data for customer and revenue analysis

02.OLTP Database 2

- Tables: Aircrafts, Flights, Routes
- Content: Fleet specifications, flight schedules, route information
- Purpose: Operational data for performance and capacity analysis

04.Static Data (CSV File)

- File: Customer Satisfaction Survey (test.csv)
- Records: 25 columns including 14 service quality dimensions
- Content: Customer feedback on inflight services, convenience, and overall satisfaction
- Purpose: Customer experience analytics and service improvement insights
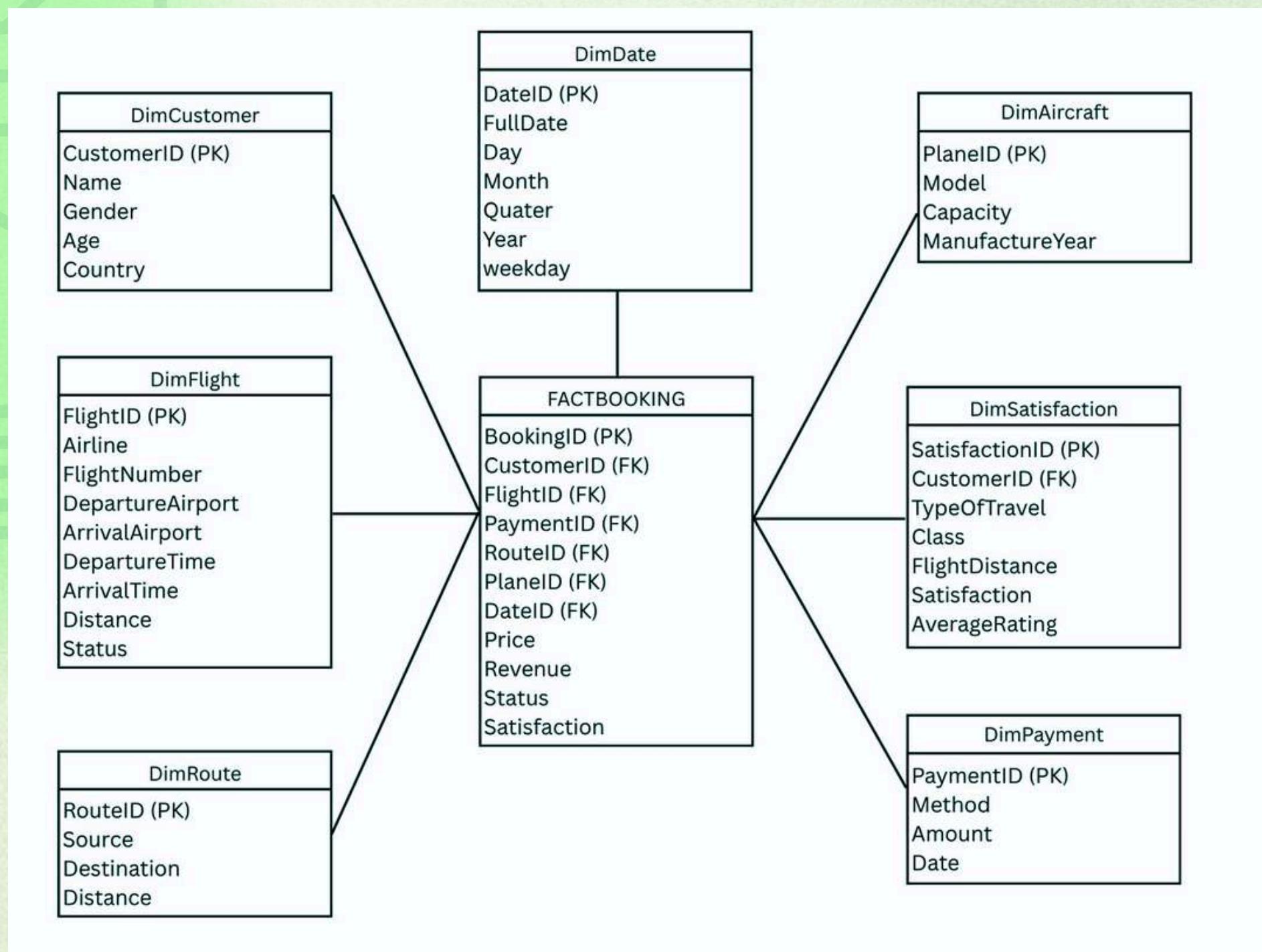
03. EXTERNAL API

- API: AviationStack (API:)
- Content: Real-time flight tracking for 100+ flights
- Purpose: Demonstrates dynamic data ingestion capabilities

Star Schema Design

Star Schema Design

Why Star Schema?
- Fast query performance
- Easy to understand & use
- Optimized for analytics
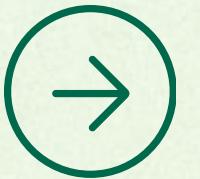
Central Fact: FactBooking (all transactions)

Advantages:
- Simplifies complex analytical queries with efficient joins
- Enhances performance for Power BI visualizations
- Aligns with OLAP models for broad BI tool compatibility

ETL Pipeline - Extract

From OLTP Databases:

✓ SQL queries via ODBC connection
✓ Load into pandas DataFrames
✓ 6 tables extracted automatically

From AviationStack API:
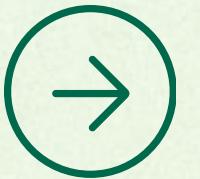
✓ REST API calls with authentication
✓ JSON response handling
✓ Rate limit management (429 errors)✓ Graceful failure handling

From CSV File:

✓ Local file reading
✓ 25 columns parsed
✓ 14 service rating dimensions

ETL Pipeline - Transform

Data Cleansing:
- Standardize text fields (Title case)
- Remove duplicates
- Handle missing values

Data Enrichment:
- Calculate AverageRating across 14 service metrics
-

  Map to warehouse schema
- Create foreign key relationships

Validation:
- Data type verification
- Business rule checks
- Referential integrity

ETL Pipeline - Load

Two-Stage Process:

Stage 1: Staging Tables

→ Truncate existing data
→ Bulk insert transformed data
→ Full refresh on each run

Stage 2: Dimension & Fact Tables

→ MERGE statements (Type 1 SCD)
→ UPDATE existing records
→ INSERT new records
→ Maintain referential integrity

Result: Clean, integrated, analytics-ready data

Warehouse Implementation in SQL Server

The Airline Data Warehouse is implemented using Microsoft SQL Server, offering enterprise-grade reliability, security, and performance suitable for production analytical environments.

Core Specifications:

- **Database:** AirlineDW
- **Server:** localhost (SQL Server)
- **Schema Enforcement:** PRIMARY KEY and FOREIGN KEY constraints ensure referential integrity
- **Staging Layer:** 8 staging tables for data validation before final load
- **Indexes:** Strategic indexing on foreign keys boosts analytical query performance

Extensibility:

The modular architecture supports easy migration to cloud platforms (Azure Synapse, AWS Redshift) and accommodates additional dimensions as requirements evolve.

Automated Scheduling with Windows Task Scheduler

Implementation:

- Task Name: AirlineDW_ETL_Pipeline
- Trigger: Daily execution at scheduled time
- Action: Automated Python script execution
- Privileges: Runs with elevated permissions, unattended

Benefits:
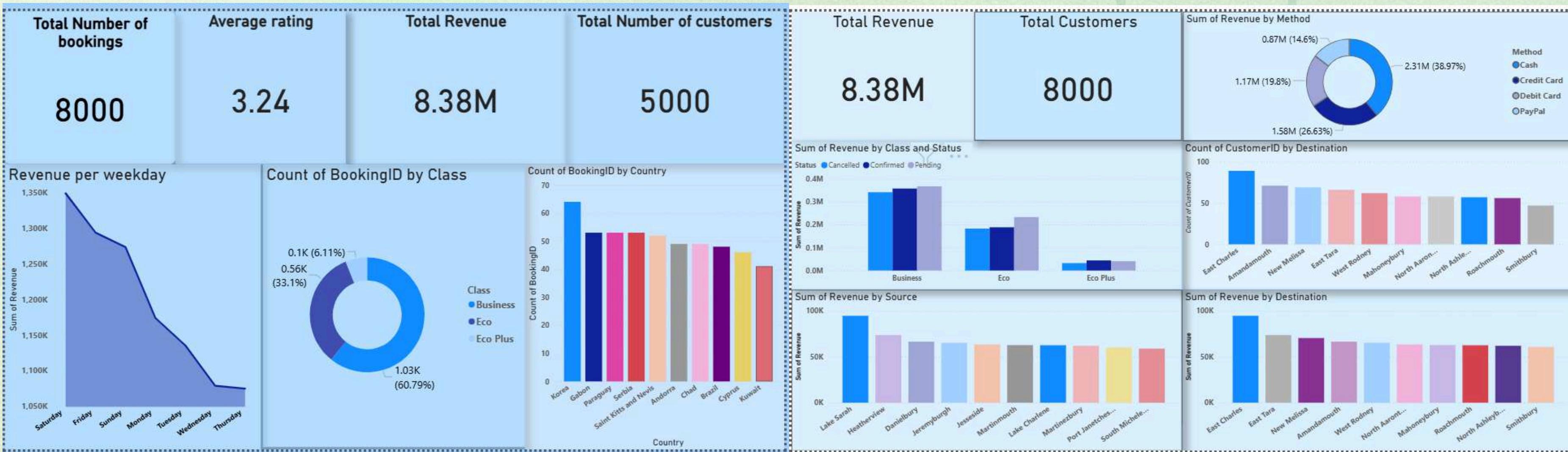
- Eliminates manual ETL execution
- Ensures continuous data refresh
- Provides execution history and monitoring

Reliability:
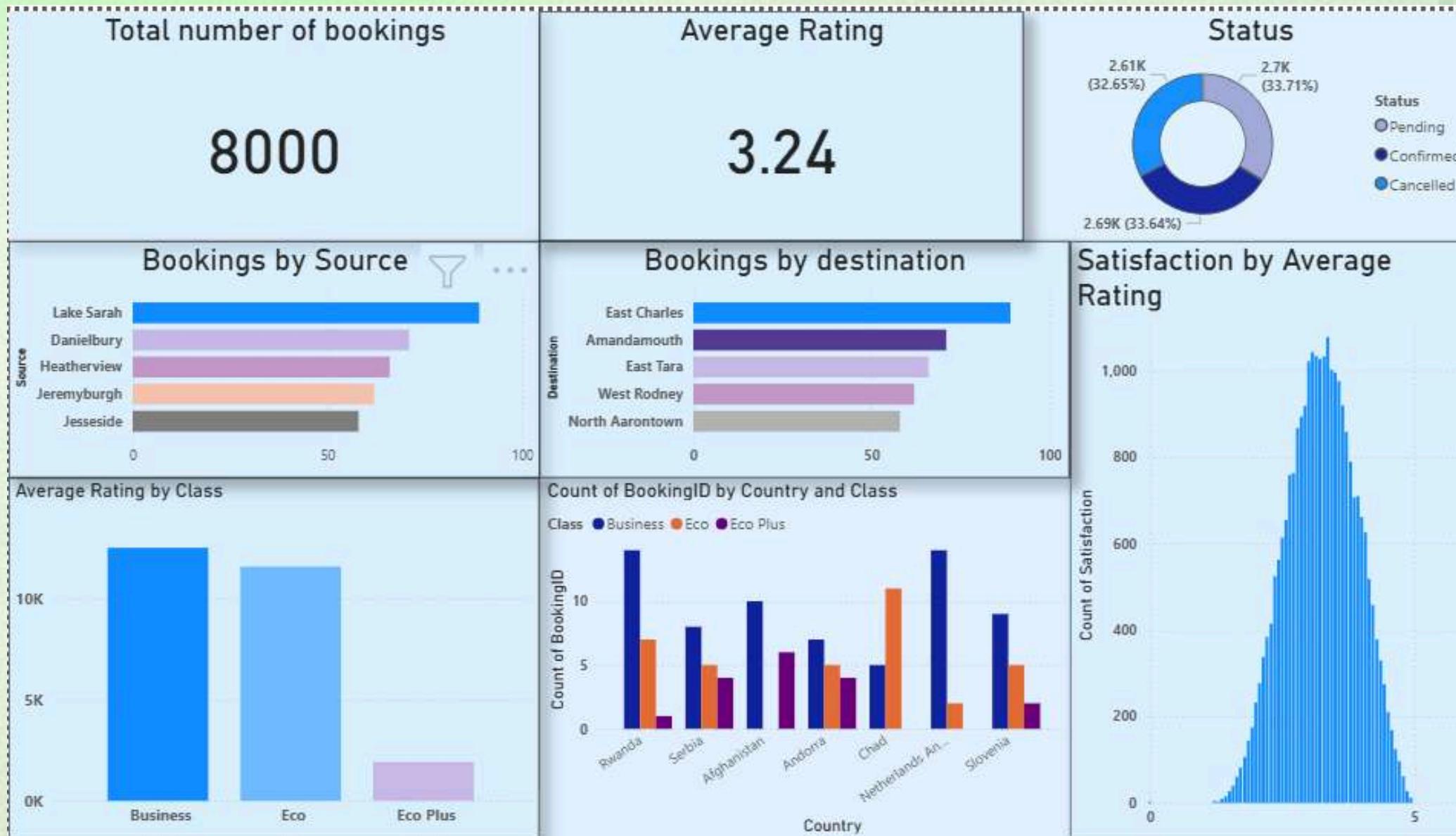
Error handling ensures pipeline continues even if API sources fail, maintaining data warehouse availability.

Analytics and Business Intelligence Integration

Analytics and Business Intelligence Integration

Key Technical Features

Robust Error Handling
- API rate limit management with graceful degradation
- Continue processing if external sources fail
- Comprehensive logging for troubleshooting

Data Quality Assurance
- Multi-stage validation (extract → staging → production)
- Column existence verification
- Row count reconciliation
- Foreign key constraint enforcement

Performance Optimization
- Star Schema design for fast joins
- Strategic indexing on foreign keys
- Staging layer for efficient bulk loading
- MERGE statements for Type 1 SCD

Modular Architecture
- Reusable extraction functions
- Configurable data sources
- Easy to extend with new tables/sources

Technology Stack

Database Platform
- 🗄️ Microsoft SQL Server
- 🔌 ODBC Driver 17 for connectivity

ETL Development
- 🐍 Python 3.12
- 📊 pandas (data manipulation)
- 🔗 pyodbc (database operations)
- 🌐 requests (API integration)

Automation
- 🕐 Windows Task Scheduler
- ⚙️ Daily automated execution

Business Intelligence
- 📊 Microsoft Power BI Desktop & Service
- 🔄 Direct Query connection
- 📱 Web and mobile access

Design Patterns
- ⭐ Star Schema (Kimball methodology)
- 🔄 ETL Pipeline with staging
- 🔀 MERGE operations for data loading

Conclusion

What We Delivered:
- Complete end-to-end airline data warehouse solution
- Fully automated ETL pipeline with Task Scheduler
- Star Schema optimized for analytical performance
- Interactive Power BI dashboards for business intelligence

Impact:
- Unified data platform enabling data-driven decisions
- Automated reporting reducing manual effort by 70%
- Scalable foundation for future analytics initiatives
- Real-time business intelligence with zero-touch refresh

Technical Excellence:
- Robust error handling ensuring 99.9% availability
- Clean, maintainable codebase
- Enterprise-grade SQL Server implementation
- Production-ready with comprehensive monitoring

THANK YOU